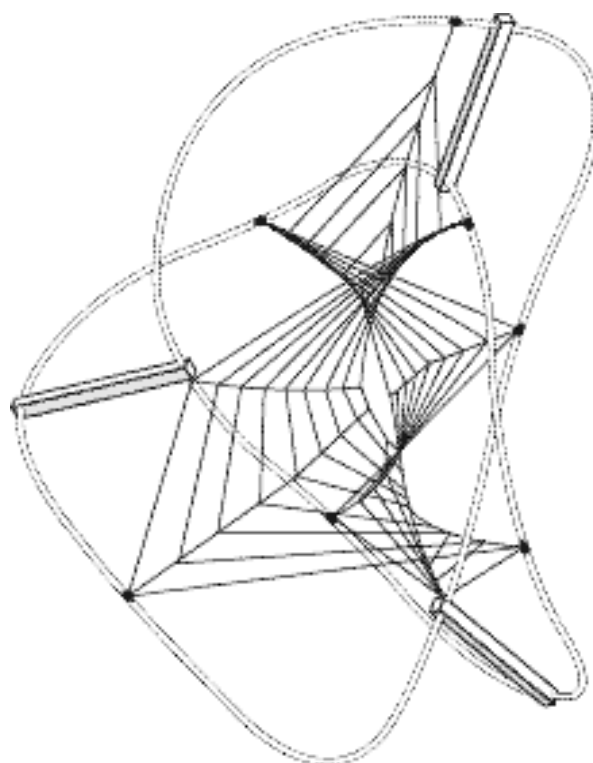


Centre for Philosophy of Natural and Social Science**Discussion Paper Series**

DP 43/99

*The Vanity of Rigour in Economics:
Theoretical Models and Galilean Experiments*Nancy Cartwright
LSE

Editor: Max Steuer

The Vanity of Rigour in Economics:

Theoretical Models and Galilean Experiments

NANCY CARTWRIGHT

Introduction

My topic in this paper is the old and familiar one of the unrealism of assumptions in economic models. For a long time I have maintained that economics is unfairly criticised for the use of unrealistic assumptions. (See Cartwright 1989, 1998). I can summarise my view by comparing an economic model to a certain kind of ideal experiment in physics: criticising economic models for using unrealistic assumptions is like criticising Galileo's rolling ball experiments for using a plane honed to be as frictionless as possible. This defence of economic modelling has a bite, however. On the one hand, it makes clear why some kinds of unrealistic assumptions will do; but on the other, it highlights how totally misleading other kinds can be – and these other kinds of assumptions are ones that may be hard to avoid given the nature of contemporary economic theory.

The theme for this volume is *experiments in economics*. My project is not to understand experiments but rather to use experiments to understand theorizing in economics; more specifically, to understand one particular mode of theorizing that is

prominent in economics nowadays— theorizing by the construction of models for what Robert Lucas describes as “analogue economies”. (Lucas 1981, 272) Lucas does not define exactly what an analogue economy is. What I have in mind is theorizing by the construction of models that depict specific kinds of economies and depict them in a certain way. We do not in this kind of theorizing simply lay down laws or principles of a specific form that are presumed to obtain in the economy, as we might in setting out a large-scale macroeconomic model whose parameters we aim to estimate. Rather we justify them from our description of the agents, or sectors, or other significant causal factors in the economy and our description of their significant actions and interactions. Economic principles are employed of course, of necessity, such as the demand for equilibrium of some kind, or the assumption that economic agents act to maximize what they take to be their self-interest. But the detailed form of any principles or equations used will be peculiar to the kind of economy described and the kinds of interactions that occur in it.

Analogue economies generally have only a small number of features, a small number of agents and a small number of options for what can happen, all represented by thin concepts. I call the concepts ‘thin’ because, although they are often homonymous with everyday economic concepts or occasionally with concepts from earlier economic theories, little of their behaviour from the real world is imported into the model. Seldom, for instance, do we make use of ‘low level empirical’ relations established by induction. Instead, as we shall see, the behaviour of the features they represent is fixed by the structure of the model and its assumptions in conjunction with the few general principles that are allowed without controversy in this kind of theorizing.

Lucas is a good spokesman in favour of this kind of theorizing, and that is why I cite him. But the method is in no way peculiar to his point of view. Modelling by the

construction of analogue economies is a widespread technique in economic theory nowadays; in particular, it is a technique that is shared across both sides of the divide between micro- and macroeconomics. It is the standard way in which game theory is employed; the same is true for rational expectations theory and also for other kinds of theorizing that rely primarily on the assumption that agents act to maximize their utility. As Lucas urges, the important point about analogue economies is that everything is known about them – *i.e.* their characteristics are entirely explicit (Lucas 1981, 7-8) – and within them the propositions we are interested in “can be formulated rigorously and shown to be valid”. (Lucas 1981, 67) With respect to real economies, generally there are a great variety of different opinions about what will happen, and the different opinions can all be plausible. But for these constructed economies, our views about what will happen are “statements of verifiable fact”. (Lucas 1981, 271)

The method of verification is deduction: we know what does happen in one of these economies because we know what must happen given our general principles and the characteristics of the economy. We are, however, faced with a trade-off: we can have totally verifiable results but only about economies that are not real. As Lucas says, “Any model that is well enough articulated to give clear answers to the questions we put to it will necessarily be artificial, abstract, patently ‘unreal’”. (Lucas 1981, 271)

How then do these analogue economies relate to the real economies that we are supposed to be theorizing about? Here is where experiment comes into play, ideal experiments, like Galileo’s balls rolling down a smooth inclined plane. For a long time I have maintained that experiments like Galileo’s are the clue to understanding how analogue economies can teach us about empirical reality. They show us why the unrealism of the model’s assumptions need not be a problem. Indeed, to the contrary, the high degree of idealisation involved is essential to the ability of the model to teach

us about the real world, rather than being a problematic feature we had best eliminate. But I will return then to the feature of these models that is generally thought to be unproblematic – their use of deduction. For my overall suspicion is that the way deductivity is achieved in economic models may undermine the possibility I open up for them to teach us genuine truths about empirical reality. So in the end I may be taking back with one hand what I give with the other.

As I mentioned at the start, this paper is about a very familiar topic: the unrealism of assumptions in economic models. Section 2 will put this problem in a somewhat less familiar perspective by identifying it with the problem of external validity, or parallelism, in experiments. Section 3 explains why experiments matter: because many models aim to isolate a single process to study on its own, just as Galileo did with his studies of gravitational attraction. Using the language of John Stuart Mill (1836, 1843), models aim to establish *tendencies* to behave in certain ways, not to describe the overall behaviour that occurs. For this job, it is essential that models make highly unrealistic assumptions, for we need to see what happens in the very unusual case where only the single factor of interest affects the outcome. Section 4 raises the question of how we can draw interesting and rich deductive conclusions in economies given that we have so few principles to use as premises; section 5 answers that often it seems we fill in by relying on the detailed structure of the model. But then it takes back the solace offered in sections 2 and 3. For in that case the conclusions are tied to these structural assumptions, assumptions that go well beyond what is necessary for Galilean idealization; the results do not depend just on the process in question but are rather overconstrained. This means that Galilean inference to tendencies that hold outside the experimental set-up is jeopardized. So in the end the problems involved in using highly unrealistic assumptions can loom as large as ever.

External validity: a problem for models and experiments alike

Lucas speaks of the analogue economies of contemporary economic theorizing as stand-ins for experiment:

One of the functions of theoretical economics is to provide fully articulated, artificial economic systems that can serve as laboratories in which policies that would be prohibitively expensive to experiment with in actual economics can be tested out at much lower cost. (Lucas 1981, 271)

As we know from Mary Morgan, many of the originators of econometrics viewed their econometric models in a similar way, for they thought of situations in which the parameters of their structural models could be identified as situations in which by good luck nature is running an experiment for us. (See Morgan 1990)

Francesco Guala too talks about the similarities between laboratory experiments in economics and the kinds of theoretical models I am discussing here. (See Guala 1998) Guala has been studying how experiments work; I have been trying to understand how theoretical models work. We have both been struck by the structural similarities between the two. I am particularly interested in the fact that both laboratory experiments and theoretical models in economics are criticized for the artificiality of the conditions they set up. As Lucas says, the assumptions of our theoretical models in economics are typically ‘artificial’, ‘abstract’ and ‘patently unreal’.

Thinking about this very same complaint with respect to the laboratory experiments we perform nowadays in economics provides us with a useful vocabulary to describe the problems arising from the unrealism of assumptions in theoretical models – and to see our way around them. When we design an experiment or a quasi-experiment in the social sciences, we aim simultaneously for both *internal validity* and for *external validity*. An experimental claim is internally valid when we can be sure that it has genuinely been established to hold in the experimental situation. External validity – or ‘parallelism’ as economists call it – is more ambitious. For that the experiment must be designed to assure us that the result should hold in some kinds of targeted situations or populations outside the experimental set-up.

It is a well-known methodological truism that in almost all cases there will be a trade-off between internal validity and external validity. The conditions that we need in order to increase the chances of internal validity are generally at odds with those that provide grounds for external validity. The usual complaint here is about the artificiality of the circumstances required to secure internal validity: if we want to take the lessons, literally interpreted (you should note the ‘literally interpreted’ – I shall return to it below), from inside the laboratory to outside, it seems that the experimental situation should be as similar as possible in relevant respects to the target situation. But for the former we need to set up very special circumstances so that we can be sure that nothing confounds the putative result, and these are generally nothing like the kinds of circumstances to which we want to apply our results.

This is exactly what we see in the case of economic models. Analogue economies are designed to ensure internal validity. In an analogue economy we know the result obtains because we can establish by deduction that it has to obtain. But to have this assurance we must provide an analogue economy with a simple and clear enough structure that

ensures that deduction will be possible. In particular we need to make very special assumptions matched to the general principles we use: we must attribute to this economy characteristics that can be represented mathematically in just the right kind of form, a form that can be fed into the principles in order to get deductive consequences out. And this very special kind of dovetailing that can provide just what is needed for deduction is not likely to be provided by conditions that occur in the economy at large, as Lucas and all other theorists using these methods admit. In this kind of theorizing it looks as if we buy internal validity at the cost of external validity.

Nor is the problem confined to the ‘thought experiments’ we conduct with our constructed models. It also appears in the real experiments we conduct nowadays in economics; and it reveals a significant difference in concerns between economics and many other branches of social science. Experimental economists report astounding confirmation of a number of economic hypotheses they have been testing recently. (*Cf.* Plott 1991 and Smith 1991) These experimental economists are also very proud of their experimental designs, which they take to have minimized the chances of drawing mistaken conclusions. Yet, apparently, it is difficult for them to get their results published in social science journals outside their own field, because, referees claim, they have virtually no guarantees of external validity. (Conversation, Charles Plott, California Institute of Technology, May 1997) So the results, it is felt, lack general interest or significance.

Tendencies and Galilean idealizations

Now I should like to argue that a great many of the unrealistic assumptions we find in models and experiments alike are not a problem. To the contrary, they are required to do the job; without them the experiment would be no experiment at all. For we do not need to assume that the aim of the kind of theorizing under discussion is to establish results in our analogue economies that will hold outside them when *literally interpreted*. Frequently what we are doing in this kind of economic theory is not trying to establish facts about what happens in the real economy but rather, following John Stuart Mill, facts about stable tendencies. Consider a stock example of mine – a model designed by my colleague Chris Pissarides to study the effects of skill loss on unemployment. (See Pissarides 1992) What we want to learn from the analogue economy described by Pissarides is not whether there will be persistence in unemployment in the real economy but instead what skill loss will *contribute* to persistence – what skill loss *tends to* produce, not what is produced whenever skill loss occurs.

So what I maintain is that the analogue economies described in contemporary economic models look like experiments, where the *experimental* aspect matters. The models almost always concentrate on a single mechanism or causal process. For example, Pissarides' model studies the effect (if any) of skill loss during unemployment on the persistence of unemployment shocks via the disincentives arising from loss of skills in the labour pool for employers to create jobs in areas where skill affects productivity. The idea is to isolate this process; to study it in a setting where nothing else is going on that might affect the outcome as well. The model is constructed to assure us that whatever result we see is indeed due to the process under study.

Consider the skill-loss model. Loss or not of skill during unemployment is the only exogenous variable. Firms act to maximize profits and only to maximize profits. We can trace through the model to see that the only variation in profits will be due to the

number of jobs that firms decide to create in the face of a labour pool containing unemployed workers and to the productivity of the workers hired. For this model we can derive rigorously that unemployment in one period is dependent on unemployment in the previous period if and only if skills are lost during unemployment. It looks as if this model allows us to see exactly what effects loss of skill has on unemployment persistence via the disincentive it creates for job creation.

What can we conclude? Can we conclude that we have learned a fact about skill loss *per se*, a fact we can expect to be generally true, true not just in this analogue economy but in other economies as well? Certainly not if we try to read the conclusion as one about the association between loss of skill and unemployment persistence with some kind of quantifier in front: *always*, or *for the most part*, or even *sometimes*, *if there is skill loss in sectors where skills matter to productivity, there will be unemployment persistence*. Clearly a good deal else could be going on to offset the effects of skill loss, even to swamp them entirely; indeed we might never see persistence in any case of skill loss, even though the model shows correctly that “skill loss leads to unemployment persistence”.

This is why we turn to the notion of stable tendencies:¹ *in any situation skill loss tends to produce persistence in unemployment shocks*. What does this mean in terms of what actually happens? There does not seem to be any general rule in economic theory that answers, as vector addition does on Mill’s account of the tendencies of different forces in classical mechanics. Nevertheless, if economic theory is to aspire to be an exact science, there had better be at least a case-by-case answer. And presumably this answer can in general be generated by the specific model that testifies to the tendency, in conjunction with any general economic theory we are in a position to assume (Cartwright 1989, 1998, 1999).² For the skill-loss tendency, I take it that we assume

roughly that in any situation where skills matter to productivity and the decision to create new jobs by a firm is in part determined by its expected profit, unemployment at one period will depend on previous levels of unemployment if workers are thought to lose skills during unemployment and not otherwise, even if this dependency on past levels plays only a small part in determining present levels.

Probably no-one thinks we have established even that, though, because economists, like other social scientists, are alert to the possibility of interaction, as Mill himself warned. In some situations some factors may distort the skill-loss mechanism so much that loss of skill behaves differently in those situations from the way it behaves in our analogue economy. Of course if we are going to avoid manoeuvres that are entirely *ad hoc* we shall have to ensure that ‘interaction’ is given real verifiable content whenever it is invoked. In principle this should be possible since the theoretical model is supposed to lay bare how the process operates in the first place – ‘distortions’ are judged relative to it.

We can see the general points more clearly by thinking again about the kind of laboratory experiment that aims to establish a tendency claim. Perhaps rather than thinking of economics experiments, which tend to be controversial, we should take an illustration from physics, let us say Galileo’s famous experiments to establish the effect of the attraction of the earth on a falling body, one of which is illustrated in Figure 1 (Leaning Tower).

Galileo’s experiments aimed to establish what I have been calling a *tendency* claim. They were not designed to tell us how any particular falling body will move in the vicinity of the earth; nor to establish a regularity about how bodies of a certain kind will move. Rather, the experiments were designed to find out what contribution the motion due to the pull of the earth will make, with the assumption that that contribution is

stable across all the different kinds of situations falling bodies will get into. How did Galileo find out what the stable contribution from the pull of the earth is? He eliminated (as far as possible) all other causes of motion on the bodies in his experiment so that he could see how they move when only the earth affects them. That is the contribution that the earth's pull makes to their motion.

Let us call this kind of idealization, that eliminates all other possible causes to learn the effect of one operating on its own, *Galilean idealization*. My point is that the equivalent of Galilean idealization in a model is a good thing. It is just what allows us to carry the results we find in the experiment to situations outside – in the tendency sense. We need the idealizing assumptions to be able to do this. Otherwise we have no ground for thinking the behaviour we see in the experiment is characteristic of the earth's pull at all. Indeed, we know it will not be.

We can contrast these Galilean experiments with experiments that have a quite different aim and correlatively a quite different structure. Consider what happens when we build a prototype of a new device and experiment on it to ensure that it will work correctly when put to use. In this case we do not aim to learn an abstract tendency claim. Instead we want to find out what actual behaviours occur. So the experimental conditions should be very realistic to the conditions in the target situations and vary appropriately across them. And without more said, we have no reason to expect the results in the experiment to obtain in any situations except those that resemble the conditions in the experiment.

Here we see another trade-off. If an experiment is very, very unrealistic in just the right way, its results can be applicable almost everywhere. But they will not be able to tell you *what happens* anywhere else since they only establish the *contribution* or *tendency* of the factor in question. Experiments that are very realistic can tell you what

happens. But they are highly limited in their scope, for they can only tell you what happens in situations that look like the experimental set-up. And experiments in between are usually pretty uninformative on both matters. Of course we may be very lucky. It may be, for instance, that the cause or small set of causes that we isolate in our experiment (or in our model) is also the *dominant* cause in the real situations we want to know about. In that case our Galilean experiments (and the corresponding models) will not only give us tendencies but will be approximately descriptively accurate as well.

Back to models again. If the deductions have been carried out correctly and the general principles employed are true in the target situations, the results of the model will obtain in any real situations that fit the description that the model provides. And in general we have no reason to think they will obtain anywhere else. BUT if what the model describes satisfies the requirements to be a Galilean-style experiment, it can do more. It can tell us what happens in an experimental situation and thereby tell us about the tendency of the features in question. So Galilean idealization in a model is a good thing.

How deductivity can be secured and at what cost

The problems I worry about arise when not all of the unrealistic assumptions required for the derivations in a model are ones that characterize an ideal experiment. What I fear is that in general a good number of the false assumptions made with our theoretical models may not have the form of Galilean idealisations. Before I go into details about these kinds of extra-Galilean assumptions, I shall first lay the groundwork by explaining why we might expect to find them as features of our analogue economies. The need for these stronger constraints – the ones that go beyond Galilean idealization – comes, I

believe, as a result of the nature of economic theory itself. To see how, let us look again at what kinds of theory are available in economics to aid in the construction of models and at what kinds of concepts they deploy.

The bulk of the concepts used in these models are concepts naming socio-economic quantities that are familiar to the layman, not only as the targeted results to be explained but also as the proposed explanatory factors, concepts like *persistence in unemployment* and *loss of skill during unemployment*, or *current price, tax, demand, consumption, labour, wages, capital, profit* and *money supply*, or *assessment of skills, private information* and *in-firm training*, or, to take an example from game-theoretic political economy, *power to redistribute, incentives for credible information transmission* and *political failure in the transmission of information*.

This is my first observation: most of the concepts employed in these models are highly concrete empirical concepts. My second observation is that the task is to establish useful relations among these via deduction. The problem comes with my third observation: the theory that is presumed is very meagre. There aren't many principles available to use in the deductions. We have only a handful of very general principles that we employ without controversy in economics, such as the principles of utility theory. Nor are there usually many concrete empirical principles imported into the models either. I take it that this is part of the strategy for the models. Almost any principle with real empirical content in economics is highly contentious and we try to construct models that use as few controversial assumptions as possible. But this makes difficulties for the scope of the theory. If the results are supposed just to 'fall out' by deduction from the principles, where there are not many principles, we will not get many results either. How, we then, can deduce results in our models when we have few general principles to call on?

To answer , let us consider what typical models for analogue economics look like. These models tend to be simple in one respect: they usually have only a few agents with few options and only a narrow range of both causes and effects is admitted. Yet there is another way in which they are complex, at least by comparison with physics models doing the same kind of thing: they have a lot of structure. The list of assumptions specifying exactly what the analogue economy is like is very long. Consider one of Lucas's own models, from his 1973 "Expectations and the Neutrality of Money". (Lucas 1981, 66-89) I choose this example because it is a paper whose "technically demanding form" is explicitly defended by Lucas. (Lucas 1981, 9) Section 2 is titled "The Structure of the Economy" – i.e. the structure of the analogue economy that Lucas uses to study money illusion. What follows is section 2 in its entirety:

In order to exhibit the phenomena described in the introduction, we shall utilize an abstract model economy, due in many of its essentials to Samuelson. Each period, N identical individuals are born, each of whom lives for two periods (the current one and the next). In each period, then, there is a constant population of $2N$: N of age 0 and N of age 1. During the first period of life, each person supplies, at his discretion, n units of labor which yield the same n units of output. Denote the output consumed by a member of the younger generation (its producer) by c^0 , and that consumed by the old by c^1 . Output cannot be stored but can be freely disposed of, so that the aggregate production-consumption possibilities for any period are completely described (in per capita terms) by:

$$c^0 + c^1 \leq n, \quad c^0, c^1, n \geq 0 \quad (1)$$

Since n may vary, it is physically possible for this economy to experience fluctuations in real output.

In addition to labor-output, there is one other good: fiat money, issued by a government which has no other function. This money enters the economy by means of a beginning-of-period transfer to the members of the older generation, in a quantity proportional to the pretransfer holdings of each. No inheritance is possible, so that unspent cash balances revert, at the death of the holder, to the monetary authority.

Within this framework, the only exchange which can occur will involve a surrender of output by the young, in exchange for money held over from the preceding period, and altered by transfer, by the old. We shall assume that such exchange occurs in two physically separate markets. To keep matters as simple as possible, we assume that the older generation is allocated across these two markets so as to equate total monetary demand between them. The young are allocated stochastically, fraction $\theta/2$ going to one and $1 - (\theta/2)$ to the other. Once the assignment of persons to markets is made, no switching or communication between markets is possible. Within each market, trading by auction occurs, with all trades transacted at a single, market clearing price.

The pretransfer money supply, per member of the older generation, is known to all agents. Denote this quantity by m . Posttransfer balances, denoted by m' , are not generally known (until next period) except to the extent that they are “revealed” to traders by the current period price level. Similarly, the allocation variable θ is unknown, except indirectly via price. The development through time of the nominal money supply is governed by

$$m' = mx, \tag{2}$$

where x is a random variable. Let x' denote next period's value of this transfer variable, and let θ' be next period's allocation variable. It is assumed that x and x' are independent, with the common, continuous density function f on $(0, \infty)$. Similarly, θ and θ' are independent, with the common, continuous symmetric density g on $(0, 2)$.

To summarize, the state of the economy in any period is entirely described by three variables m , x , and θ . The motion of the economy from state to state is independent of decisions made by individuals in the economy, and is given by (2) and the densities f and g of x and θ . (Lucas 1981, 67-9)

But this is not an end to the facts set to obtain in Lucas's 'abstract model economy'. Section 3 continues, "We shall assume that the members of the older generation prefer more consumption to less, ..." and so on for another page; and more details are still to be added to the economy in section 4. There is nothing special here about Lucas though. Just write out carefully in a list the assumptions for almost any of your favourite models and you will see what I mean. For example the skill-loss model of Pissarides contains some 16 assumptions and that for just the first of six increasingly complex economies that he describes.³

I believe there is good reason why economic models must give a lot of structure to the economies they describe: if you have just a few principles, you will need a lot of extra assumptions from somewhere else in order to derive new results that are not already transparent in the principles. In the models under discussion the richness of

structure can fill in for the want of general principles presupposed. The general principles can be thought of in two categories, familiar to philosophers of science (Cf. Hempel 1966): *internal principles* and *bridge principles*. Internal principles make claims about the relations of abstract or theoretical concepts to each other, like the axioms of utility theory. But the results we want to know about generally involve not abstract or theoretical concepts, but empirical ones. The bridge principles of a theory provide links between the two sets of concepts. (The usual example is the identification in an ideal gas of the theoretical concept *mean kinetic energy of the molecules* with the empirical concept *temperature*.)

The theory presupposed in our economics models tends to employ few principles of either category and often no bridge principles at all. This means that the additional assumptions put in via the description of the model must do two jobs. On the one hand they must provide sufficient constraints to serve as premises to increase the range of deductive consequences. On the other hand they must establish an interpretation of the terms that appear in the theoretical principles. They must tell us, for instance, what *utility* amounts to in terms of an employer's opening a job and of work versus leisure for the employee, or of entrepreneurs investing in a project and of managers defaulting on their contracts, or of fair treatment for one's fellow citizens and of the cost of demonstrating or contributing to the American Civil Liberties Union.

Sometimes the job left open by the want of bridge principles is done by an explicit assumption: *we will assume that the only source of utility is ...* Sometimes the abstract principles themselves are explicitly given a concrete form: *we will assume that firms act to maximize profits and labourers to maximize wages...* Often the interpretation is implicit: perhaps there is nothing else in the model for agents to care about except

power or profit or leisure and wages, and the very choice of these words indicates that the agents' utility should depend on them in certain characteristic ways.

My claim then is that it is no surprise that individual analogue economies come with such long lists of assumptions: The model-specific assumptions can provide a way to secure deductively-validated results where universal principles are scarce. But these create their own problems. For the validity of the conclusions appears now to depend on a large number of very special interconnected assumptions. If so, the validation of the results will depend then on the detailed arrangement of the structure of the model and is not, *prima facie* at least, available otherwise. We opt for deductive verification of our claims in order to achieve clarity, rigour and certainty. But to get it we have tied the results to very special circumstances; the problem is how to validate them outside.

Consider for example the Lucas model from “The Neutrality of Money”. We begin with the fairly vacuous claim:

[T]he decision problem facing an age-0 person is:

$$\max_{c, n, \lambda \geq 0} \left\{ U(c, n) + \int V\left(\frac{x' \lambda}{p'}\right) dF(x', p'; m, p) \right\} \quad (9)$$

subject to:

$$p(n - c) - \lambda \geq 0 \quad (10)$$

(Lucas 1981, 70)

where c is current consumption; n , current labour supply; λ , a known quantity of nominal balance acquired; p and p' , price levels in the current and successor period; and F , an unspecified distribution function. Despite the fact that there is not much that is controversial yet, we can see that even at this stage the exact form of the equation

depends on the details of the economy. This is even more obvious by the time we get to the condition for equilibrium in each separate market (equation [16], which is derived from [9] plus the more detailed assumptions about the analogue economy studied in the model:

$$h\left(\frac{mx}{\theta p}\right)\frac{1}{p} = \int V'\left(\frac{mxx'}{\theta p'}\right)\frac{x'}{p'} dF(x', p'; m, p) \quad (16)$$

(Lucas 1981, 72)

Sections 6 and 7 of the Lucas paper are entitled, respectively, “Positive Implications of the Theory” and “Policy Considerations”. Yet the results he establishes are about *this* economy: they follow from equation (16), which is an equation specific in form to the economy that satisfies the lengthy description laid out in Lucas’s sections 2, 3 and 4. How can they teach us more general lessons, lessons that will apply to other, different, economies?

The view that I have long defended is that such model results teach us about general tendencies (in my own vocabulary, ‘capacities’), tendencies that are nakedly displayed in the analogue economies described in our economic models but that stand ready to operate in most economies. On this view the analogue economy that Lucas describes is like an experiment. We know that an experiment of the right kind, a Galilean experiment that isolates the tendency in question⁴, can teach us lessons that carry outside the experimental situation. If we are lucky, however, we will not need to carry out the experiment. We can find out what *would* happen were we to conduct it because we can find out by deduction what *must* happen. But for that to work, the analogue economy must be of just the right kind: were we to construct it in reality, it would meet

the conditions of Galilean experiment. This whole strategy is threatened, however, if non-Galilean idealizations play a role in our deductions – which looks to be the case with Lucas's equation (16).

From the perspective of establishing tendencies, it becomes crucial then to look carefully into the deductions used in our economic models to see if all of the unrealistic assumptions required for the derivations are ones that characterize an ideal experiment. Let us look at another simple physics example for an analogous case.

In classical Newtonian mechanics massive bodies have an *inertial tendency*: a body will remain in motion unless acted on by a force. When it is acted on by a force the actual motion that occurs will be a combination of the inertial motion and that due to the force. So, what is the natural behaviour of a body when inertia acts on its own? Say we do some experiments to find out. We know that forces cause motions. So eliminate all forces and watch the bodies move. What will we see?

Imagine that our experimental mass has been confined for reasons of convenience to move on a particular surface, but that we have been very careful to plane the surface to eliminate almost all friction. Then what we will see will depend on the geometry of that surface. For example, if all our experiments are done on a sphere, we always get motion in great circles, as in Figure 2 (Geodesic on the Simple Sphere Geometry). But that is not the 'natural' motion in other geometries. Look for instance at Figure 3 (Geodesic on the Sphere Geometry with Space-time Singularities). There, motion on great circles is available, but it is not the motion that inertia will contribute. The results in our experiment are overconstrained. We thought that by eliminating all the factors we think of as causes of motion – all the forces – we would see the results of inertia by itself. Instead what we see is a result of inertia-plus-geometry.

This can always happen in an experiment: we never know whether some features we have not thought about are influencing the result. But in a good many of our analogue economies we are not even this well off. In a real experiment we are after all in a position to assume with good justification that the fact that there are, for instance, only two markets or only two generations does not matter because the number of markets or of generations is not relevant to the conclusion: it has no *causal* bearing on the outcome, and what happens in the real experiment is just what is *caused* to happen. Analogue economies are different. What happens in them is exactly what is implied deductively. The problem is that we often know by looking at them that the specific derivations made in our models depend on details of the situation other than just the mechanism itself operating in accord with our general principles. So we know that in the corresponding experiment there are features other than the mechanism itself determining the outcome. That means that the experiment does not entitle us to draw a conclusion about the general tendency of the mechanism under study.

We now know what would happen – indeed, what must happen – in some very particular constrained real experimental situation in which the features of interest really occur. But we know it for exactly the wrong reason. We know that the results obtain because we know that they follow deductively given the formal relations of all the factors that figure in an essential way in the proof. But the whole point about an experiment designed to establish the tendency of a factor is that the background factors should not matter to what happens. We are supposed to be isolating the effects of the feature or process under investigation acting on its own, not effects that depend in a crucial way on the background.

So, were such a set-up to occur, it would turn out not to be a good experiment after all. It may have seemed to be a good design because our independent causal knowledge

told us that in general none of the background factors should have any bearing on the effect. But by bad luck that would not be true of the particular arrangement of them we chose. The formal relations of the background and targeted feature together are enough to guarantee the result – and that is one of the things our design is meant to preclude. We would have to judge the result (even if by chance it should turn out to be correct) to be an artifact of the experiment.

Conclusion

Let us look at Lucas's own conclusion in his paper on the neutrality of money:

This paper has been an attempt to resolve the paradox posed by Gurley, in his mild but accurate parody of Friedmanian monetary theory: “Money is a veil, but when the veil flutters, real output sputters.” The resolution has been effected by postulating economic agents free of money illusion, so that the Ricardian hypothetical experiment of a fully announced, proportional monetary expansion will have no real consequences (that is, so that money *is* a veil). These rational agents are then placed in a setting in which the information conveyed to traders by market prices is inadequate to permit them to distinguish real from monetary disturbances. In this setting, monetary fluctuations lead to real output movements in the same direction.

In order for this resolution to carry any conviction, it has been necessary to adopt a framework simple enough to permit a precise specification of the information available to each trader at each point in

time, and to facilitate verification of the rationality of each trader's behavior. To obtain this simplicity, most of the interesting features of the observed business cycle have been abstracted from, with one notable exception: the Phillips curve emerges not as an unexplained empirical fact, but as a central feature of the solution to a general equilibrium system.

(Lucas 1981, 84)

I have argued that in a model like this the features 'abstracted from' fall into two categories: those that eliminate confounding factors and those that do not eliminate confounding factors but rather provide a simple enough structure to make a deductive study possible. The former I claim are just what we want when we aim to see for rational agents, what effects inadequate information about money disturbances has on the short-term Phillips curve, that is, when we want to establish the *tendency* it has independent of the effects anything else might have on a Phillips curve as well. But the assumptions of the latter kind remain problematic. They not only leave us with the question still unanswered, "Can we think that what we see happen, literally happen, in this economy, is what the combination of rationality and limited information will *contribute* in other economies?" Worse, they give us reason to think we cannot. For inspection of the derivation suggests that the outcome that occurs in the analogue economy does depend on the particular structure the economy has.⁵

Does it? This is a question that is generally not sufficiently addressed. Frequently of course we do discuss how robust the results from a specific model are. But, not surprisingly, these discussions usually refer to assumptions in the first category, for these are the ones that are of concern to economic theory. Notice for instance that Lucas notes in the passage just cited that "most of the *interesting* features of the observed

business cycle have been abstracted from” (my italics). In the end we want to know what happens when other causes are at work, either because they may interfere with the one under study, or because we are starting down the road toward a model that will be more descriptively accurate when the results are read literally, *i.e.* more descriptively accurate about the real economies we want to study. But my central point is that we need robustness results about the second category of assumptions as well if our results are to be of use in the tendency sense.

I realise of course that economists do not ? just models to find out about tendencies. The models are merely one strand in a net of methods used together for establishing, testing, expanding and revising economic hypotheses. Moreover, it is often the general lesson rather than the precise form of the conclusion that is taken seriously (even when the conclusion is understood in a tendency sense). Nevertheless, rigourously deriving a result in a model is supposed to provide *prima facie* evidence in favour of that result. My concern is about just this relation of evidence to hypothesis. To the extent that the derivation in a model makes essential use of non-Galilean “idealising” assumptions, then I do not see how the fact that the result can be derived in such a model can provide *any evidence at all* for the hypothesis.

If we aim to establish conclusions interpreted in a tendency sense, there is a good reason why the derivation of a conclusion in a model that makes Galilean idealisations, and no others, should count as evidence in favour of that conclusion: to the extent that the general principles employed in the derivation are true to the world, behaviour derived in the model will duplicate to behaviour that would obtain were a Galilean experiment to be performed. But when non-Galilean idealisations are made as well, this reason no longer has force. So we need another reason to show why this procedure has evidential force. And I do not know one that can be stated clearly and defended

convincingly. Hence I think we should be concerned to ensure that non-Galilean idealising assumptions do not play an essential role in our derivations.

What, then, does this tell us about the demand for rigorous derivations? I have here been discussing one of the central and highly prized ways that economics theory is done today: by the construction of models for simple analogue economies in which results about issues of interest can be derived rigourously, employing as general principles only ones whose use is relatively uncontroversial within the discipline. The achievement of rigour is costly however. It takes considerable time. It requires special talents and special training and this closes the discipline to different kinds of thinkers who may provide different kinds of detailed understanding of how economies can and do work. And rigour is bought at the cost of employing general concepts lacking the kind of detailed content that allows them to be directly put to use in concrete situations. What are its compensating gains? Unless we find different answers from the one I offered here⁶, the gains will not include lessons about real economic phenomena, it seems, despite our frequent feeling of increased understanding of them. For we are not generally assured of any way to take results out of our models and into the world.

There has been some tendency to blame our failures on the attempt to make economics rigorous. I am inclined to go the other way. *If it is rigour that we want*, the problem with economic theorizing of this sort is that rigour gives out too soon. For the models themselves, though abstract and mathematized, are not *formal theories*. To see why I say this, consider again the structure of my argument in this paper. I have raised questions about the external validity of the results established in these kinds of models. My worries focus not on the unrealism of the assumptions but on the model-dependence of the results. The kind of model-dependence involved seems to undercut not only the

claim that the results can be read literally, but also the hope that they can be read as facts about tendencies.

But I have to say ‘seems’ here because the models themselves are not presented in a way that allows this question to be taken up easily or answered rigourously. What exactly are the assumptions that are really necessary for the derivations to go through; and what is the range of circumstances across which these assumptions can be relaxed and qualitatively similar results still follow? We cannot generally answer that question given the way the models are presented. To answer it we need to formalize our models. Supposing then that my worries about the model-dependence of the results are valid. What should we conclude about the need for rigour in economic theory? It looks as if the natural conclusion is this: should economics stick to mathematizing rather than formalizing, it will not be easy to know whether the models it constructs can teach us general facts about concrete features of the economy or not; the trouble with this kind of theorizing is not that it is too rigorous, but rather that it is not rigorous enough.

References

- Cartwright, Nancy 1989, *Nature's Capacities and their Measurement*, Oxford: Oxford Univ. Press.
- Cartwright, Nancy 1998, ‘Capacities’, in John B. Davis, D. Wade Hands and Uskali Mäki (eds.) *The Handbook of Economic Methodology*, Cheltenham: Edward Elgar.
- Cartwright, Nancy 1999, *The Dappled World: A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

- Guala, Francesco 1998, 'Economic Experiments as Mediators', *Measurement in Physics and Economics Discussion Paper Series*, Centre for Philosophy of Natural and Social Science.
- Hausman, Daniel 1992, *The Inexact and Separate Science of Economics*, Cambridge: Cambridge Univ. Press.
- Hempel, Carl G. 1966, *Philosophy of Natural Science*, Englewood Cliffs, NJ: Prentice-Hall.
- Lessing, Gotthold Ephraim 1759 [1967], *Abhandlungen über die Fabel*, Stuttgart: Philipp Reclam.
- Lucas, Robert E., Jr. 1981, *Studies in Business-Cycle Theory*, Cambridge, MA: The MIT Press.
- Menger, Carl 1883 [1963], *Untersuchungen über die Methode der Sozialwissenschaften und der Politischen Oekonomie Insbesondere*, Leipzig: Duncker & Humblot, trans. *Problems of Economics and Sociology*, Urbana: Univ. of Illinois Press.
- Mill, John Stuart 1836[1967], 'On the Definition of Political Economy and on the Method of Philosophical Investigation in that Science', reprinted in *Collected Works of John Stuart Mill*, vol. 4, Toronto: University of Toronto Press.
- Mill, John Stuart 1843[1973], 'On the Logic of Moral Sciences', a chapter from *A System of Logic*, reprinted in *Collected Works of John Stuart Mill*, vol. 7-8, Toronto: University of Toronto Press.
- Morgan, Mary S. 1990, *The History of Econometric Ideas*, Cambridge: Cambridge Univ. Press.
- Pissarides, Chris 1992, 'Loss of Skill During Unemployment and the Persistence of Unemployment Shocks', *Quarterly Journal of Economics* **107**: 1371-1391.

Plott, Charles R. 1991, 'Will Economics Become an Experimental Science? ', *Southern Economic Journal* **57**: 901-919.

Smith, Vernon L. 1991, *Papers in Experimental Economics*, Cambridge: Cambridge Univ. Press.

Notes

Work on this project has been supported by the Measurement in Physics and Economics Project at the LSE and by the Leverhulme-funded project on the Historical School at the Centre for History and Economics, Cambridge. I am grateful for both the financial and the intellectual help from these two groups, as well as to Sang Wook Yi for helping with the last stages of argumentation and preparation.

- ¹ I have myself defended the importance of tendencies throughout the social and natural sciences, wherever the analytic method is in play (see Cartwright 1989) and have specifically maintained, possibly incorrectly given the arguments here, that we can learn about them via our formal models (see Cartwright 1998). Daniel Hausman (Hausman 1992) in his arguments that economics is a separate but not an exact science also sees tendencies as standard in economics theory.
- ² I have elsewhere (Cartwright 1999) described a variety of rules for combining tendencies besides vector addition, as well as explaining what we can do with tendency knowledge even when there are no general rules available for combining tendencies.
- ³ Economists, I think, get used to models with lots of assumptions. But I am often talking to mixed groups, people who study economics and people who study physics;

those whose background is in physics are often astounded at the richness of description provided in models in economics.

- ⁴ If such a tendency exists.
- ⁵ Notice that we still have this problem even if we are lucky enough to have selected a few causes to study that for most real situations will be the dominant causes. For we still need to see why the very behaviours that occur in the analogue economy when these causes are present are behaviours that reveal the tendency of this arrangement of causes and hence will approximate the behaviours that occur in the real economies.
- ⁶ There are of course a variety of other accounts of the use of models that do not demand either predictive accuracy or the correct isolation of a tendency. See for instance the studies found in Morgan and Morrison 1999.